




## Article

# VetCompass Australia: A National Big Data Collection System for Veterinary Science

Paul McGreevy <sup>1</sup>, Peter Thomson <sup>2</sup>, Navneet K. Dhand <sup>1</sup>, David Raubenheimer <sup>3</sup>,  
Sophie Masters <sup>1,\*</sup>, Caroline S. Mansfield <sup>4</sup>, Timothy Baldwin <sup>5</sup>, Ricardo J. Soares Magalhaes <sup>6,7</sup>,  
Jacquie Rand <sup>6</sup>, Peter Hill <sup>8</sup>, Anne Peaston <sup>8</sup> , James Gilkerson <sup>9</sup>, Martin Combs <sup>10</sup>,  
Shane Raidal <sup>10</sup>, Peter Irwin <sup>11</sup>, Peter Irons <sup>11</sup>, Richard Squires <sup>12</sup>, David Brodbelt <sup>13</sup> and  
Jeremy Hammond <sup>14</sup>

<sup>1</sup> Sydney School of Veterinary Science, Faculty of Science, University of Sydney, Sydney, NSW 2006, Australia; paul.mcgreevy@sydney.edu.au (P.M.); navneet.dhand@sydney.edu.au (N.K.D.)

<sup>2</sup> School of Life and Environmental Sciences, Faculty of Science, University of Sydney, Sydney, NSW 2006, Australia; peter.thomson@sydney.edu.au

<sup>3</sup> Charles Perkins Centre and School of Life and Environmental Sciences, University of Sydney, Sydney, NSW 2006, Australia; david.raubenheimer@sydney.edu.au

<sup>4</sup> Faculty of Veterinary and Agricultural Sciences, University of Melbourne, Werribee, VIC 3030, Australia; cmans@unimelb.edu.au

<sup>5</sup> School of Computing and Information Systems, University of Melbourne, Parkville, VIC 3010, Australia; tb@ldwin.net

<sup>6</sup> School of Veterinary Science, University of Queensland, Gatton, QLD 4343, Australia; r.magalhaes@uq.edu.au (R.J.S.M.); j.rand@uq.edu.au (J.R.)

<sup>7</sup> Child Health Research Centre, University of Queensland, South Brisbane, QLD 4101, Australia

<sup>8</sup> School of Animal and Veterinary Sciences, University of Adelaide, Roseworthy, SA 5371, Australia; p.hill@adelaide.edu.au (P.H.); anne.peaston@adelaide.edu.au (A.P.)

<sup>9</sup> Faculty of Veterinary and Agricultural Sciences, University of Melbourne, Parkville, VIC 3010, Australia; jrgilk@unimelb.edu.au

<sup>10</sup> School of Animal and Veterinary Science, Faculty of Science, Charles Sturt University, Wagga, NSW 2650, Australia; macombs@csu.edu.au (M.C.); shraidal@csu.edu.au (S.R.)

<sup>11</sup> School of Veterinary and Life Sciences, Murdoch University, Murdoch, WA 6150, Australia; p.irwin@murdoch.edu.au (P.I.); P.Irons@murdoch.edu.au (P.I.)

<sup>12</sup> College of Public Health, Medical and Veterinary Science, James Cook University, Townsville, QLD 4811, Australia; richard.squires@jcu.edu.au

<sup>13</sup> Pathobiology and Population Services, Royal Veterinary College, University of London, Hertfordshire AL9 7TA, UK; dbrodbelt@rvc.ac.uk

<sup>14</sup> Information and Communications Technology, University of Sydney, NSW 2006, Australia; jeremy.hammond@sydney.edu.au

\* Correspondence: sophie.masters@sydney.edu.au; Tel.: +61-409-990-037

Received: 4 August 2017; Accepted: 21 September 2017; Published: 26 September 2017

**Simple Summary:** The VetCompass Australia program collects real-time clinical records from veterinary practices and aggregates them for researchers to interrogate. It delivers Australian researchers sustainable and cost-effective access to authoritative data from hundreds of veterinary practitioners, across Australia and opens up major international collaborative opportunities with related projects in the United Kingdom and elsewhere.

**Abstract:** VetCompass Australia is veterinary medical records-based research coordinated with the global VetCompass endeavor to maximize its quality and effectiveness for Australian companion animals (cats, dogs, and horses). Bringing together all seven Australian veterinary schools, it is the first nationwide surveillance system collating clinical records on companion-animal diseases and treatments. VetCompass data service collects and aggregates real-time, clinical records for researchers to interrogate, delivering sustainable and cost-effective access to data from hundreds of

veterinary practitioners nationwide. Analysis of these clinical records will reveal geographical and temporal trends in the prevalence of inherited and acquired diseases, identify frequently prescribed treatments, revolutionize clinical auditing, help the veterinary profession to rank research priorities, and assure evidence-based companion-animal curricula in veterinary schools. VetCompass Australia will progress in three phases: (1) roll-out of the VetCompass platform to harvest Australian veterinary clinical record data; (2) development and enrichment of the coding (data-presentation) platform; and (3) creation of a world-first, real-time surveillance interface with natural language processing (NLP) technology. The first of these three phases is described in the current article. Advances in the collection and sharing of records from numerous practices will enable veterinary professionals to deliver a vastly improved level of care for companion animals that will improve their quality of life.

**Keywords:** companion animals; canine; feline; equine; disease surveillance; veterinary; electronic patient record; epidemiology; big data

---

## 1. Background

The analysis of data on clinical management of companion-animal disorders provides critical evidence on the ranking of key conditions affecting animal health and welfare, as well as an evidence base for the cost-effectiveness of competing treatments and care options for these animals. Australia has one of the highest rates of pet ownership in the world, with 38.5% of the 8 million Australian households owning at least one dog and 29.2% owning a cat [1]. There are also approximately 570,000 pleasure horses in Australia [2]. The widespread distribution of companion animals means that their health and welfare have a significant and direct impact across a broad socio-economic cross-section of the Australian population. Many areas of companion-animal research also have direct translations to and implications for human health, such as investigations into zoonoses (e.g., rabies and Hendra virus).

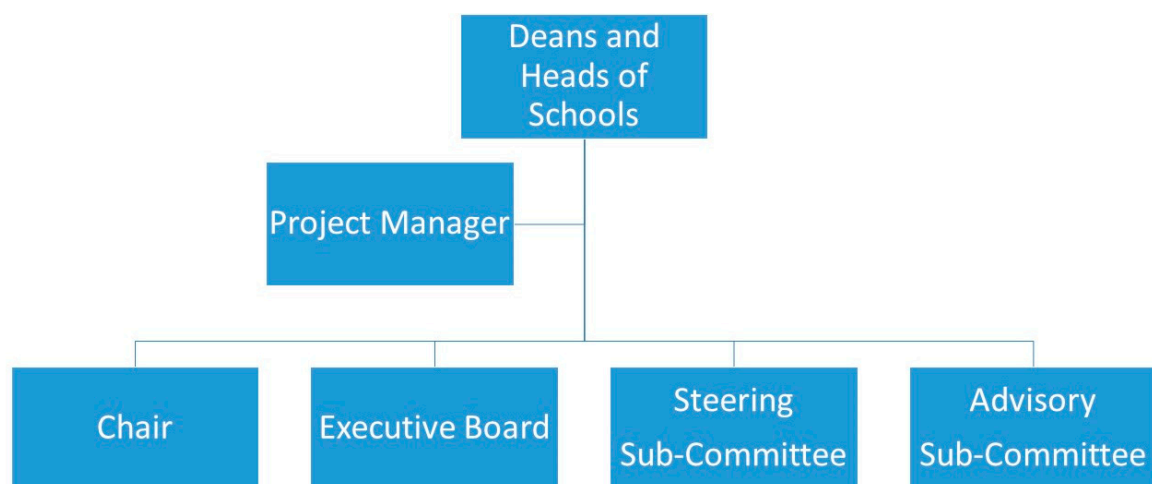
This article describes a collaborative project that builds on a similar approach adopted by the lead author and others at the Royal Veterinary College (RVC), UK. VetCompass Australia uses a software platform established at the RVC to harvest clinical records from veterinary practices into a centralized resource, enabling access to comprehensive and real-time primary practice data. It has been running in the UK for seven years, and is already transforming veterinary science there as the leading source of epidemiological data, with nearly 40 million treatment records related to approximately 6 million animals, from 498 practices available for interrogation [3,4]. In Australia, sentinel surveillance approaches have been used for a number of companion-animal diseases. The Disease WatchDog surveillance system was a prospective national disease-surveillance project, focusing on dogs and cats [5]. Data entered into the system from veterinary clinics as cases of diseases were diagnosed provided monthly online updates. However, recent analysis of its data has indicated the existence of significant data artefacts that are likely to be a result of poor compliance by users of the system [6]. The VetCompass data service overcomes these challenges in that, once implemented on practice-management systems in Australia, it allows automated transmission and de-identification of veterinary clinical records into a centralized database. Researchers can then immediately interrogate the electronic records held by participating clinics. As a long-term investment, the research potential grows as new records are entered and as new practices sign up. In addition, the project will incorporate new linguistic programming methodology to improve the quality of data extraction from those records.

The project will significantly and rapidly advance the understanding of companion-animal diseases in Australia. Bringing together all seven Australian veterinary schools, the project is based on a consortium that operates a shared resource to support the most cost-effective mechanism to undertake companion-animal veterinary research: epidemiology on aggregated electronic patient records. The Australian government has provided support through the Australian Research Council's (ARC) Linkage Infrastructure and Equipment Fund (LIEF), which will enable the comprehensive

reporting of data that are collected routinely in Australian veterinary clinics but are not currently integrated and are so far virtually untapped. This is the single most important development required to support veterinary research across a range of fields, including public health. It is expected to transform veterinary-science research and the curricula at all Australian veterinary faculties.

The project has been designed to develop across three essential phases of work: (1) roll-out of the VetCompass platform to harvest Australian veterinary clinical data; (2) continued development of the data interrogation platform to allow researchers and clinicians to realize their research potential; and (3) creation of a world-first real-time surveillance interface with natural language processing (NLP) technology. The VetCompass platform will open the way for collaborating veterinary schools to prioritize research goals and monitor the emergence and/or re-emergence of biosecurity threats that are unique to Australia as an island nation. The project is now entering Phase 2 and can be contacted via [www.vetcompass.com.au](http://www.vetcompass.com.au).

As of August 2017, the LIEF Consortium (the 15 co-investigators on the ARC grant, who represent all the Australian veterinary schools) has agreement from 180 Australian veterinary clinics to participate in the VetCompass program, and has commitments from industry and peak body partners to increase this to 80% of Australian veterinary practices. The University of Sydney's Human Research Ethics Committee has granted approval to harvest historic and future data from episodes of veterinary care (Approval number: 2013/919). The organizational framework appears in Figure 1.



**Figure 1.** A diagram to illustrate the organizational framework of VetCompass Australia.

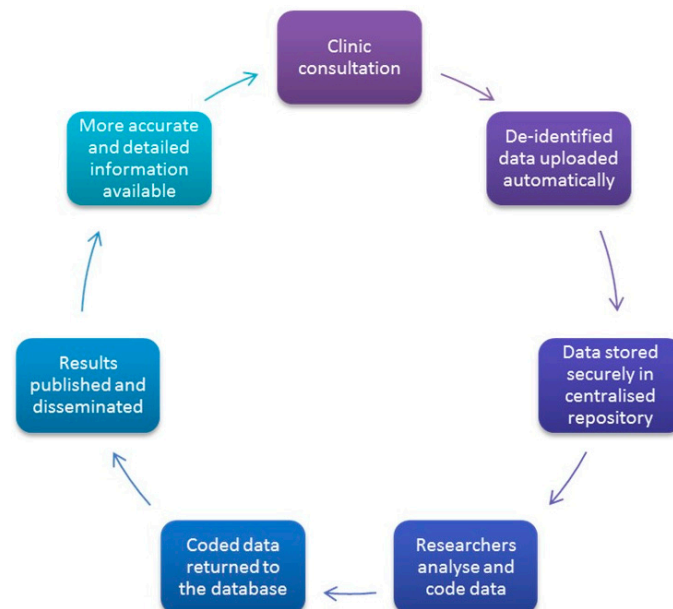
## 2. How It Works

When veterinary clinics express interest in taking part in VetCompass Australia, they are given a copy of the Participant Information Statement, which outlines the project and has HREC approval. They then complete a Practice Consent Form. They have the option to withdraw from the study at any point, but any data already transferred remains for the use of researchers.

VetCompass Australia mirrors its UK sister project. The platform ([www.VetCompass.org](http://www.VetCompass.org)) has two main components. Firstly, there are web-services with a single page application (angular) user interface for these web-services. Some of the web-services will be hosted by the University of Sydney to service requests originating in Australia, while others are hosted in and federated to the UK to service global requests. Secondly, there are back-end databases that include a Structured Query Language (SQL) server for persistence, disaster recovery, and staging data as they are imported, and a search database for searching and other denormalization of the persisted data.

VetCompass extracts data directly from practice management software (PMS) so that no extra work is required of clinicians (see Figure 2). The VetCompass operations team, in collaboration with the practice owners, implement an integrated and automated harvester on each clinic site (via application

programming interfaces). This harvester is part of the development undertaken by the project. This aligns with our commitment to observe practice without affecting care. Extracted data are uploaded directly from practices' databases via our web-services platform to the VetCompass database. The uploads vary with practices/groups but are frequently on a daily basis (e.g., overnight).



**Figure 2.** An illustration of the flow of data through the VetCompass Australia system.

Data uploaded as described above are pushed via encrypted web-services to the secure VetCompass database where they are imported into the relevant database tables. After the initial import of a dataset, these data are checked for internal inconsistencies, such as breed versus species data, and whether all records appear to have been uploaded, for example, presence of free-text notes. Data are evaluated in detail during research studies within the online interface.

The data are stored securely in the University of Sydney's servers. We have undertaken best practice information security procedures, conducting architecture reviews, and independent third-party penetration tests. The consortium owns all the data and other material contained within the database, including any data or other material entered by the researchers using the system. Aspects of the metadata are shared with the Royal Veterinary College to assist in making improvements to the platform.

All aspects of the study are treated as strictly confidential, and only researchers in VetCompass Australia or collaborating research projects, such as VetCompass at the RVC, have access to the raw data. We abide by Australian privacy legislation. Individual practices, animals, or clients are not identifiable in any publications resulting from research using this system.

Personal details of the owners are removed from the record before reaching researchers. The only exception to this is the postcode, which is collected to allow for geographical analysis of the data for certain studies. The animals themselves are not named; instead, each is given a unique identifier. This allows us to collate complete case histories from separate visits. We also collect the animal's microchip number so that its record can be traced across different clinics, specialists, and owners. This also allows us to mesh patient data with records from owners who use the smartphone app doglogbook ([www.doglogbook.com](http://www.doglogbook.com)).

Owners who do not want their animals to be a part of this project can notify clinic staff members, who then mark the record accordingly. We then remove them from data transfers and do not use their records in any studies.

With a focus on extraction and cleaning of data from numerous PMSs, the team at the RVC has developed a canonical form which offers a very simple data schema for sending data. The canonical form is implemented in various formats, such as XML Schema Definition, which allows for the selection of optimal implementation for each PMS. The development team is engaged in ongoing work with either the PMS developers or the in-house developers from practice groups to map model PMS data to the canonical form. This process facilitates the import data from the canonical form into VetCompass database. Once they have been imported, the data are subjected to automated checks for common problems and undergo a validation process. Around 60–70% of practices in Australia use RxWorks. The current software handles the extraction of data from RxWorks and is made available to in-house developers at practice groups to deploy and direct at their own databases.

The VetCompass Australia facility has its project manager and server based at the University of Sydney. There is commitment from the Australian schools to support the facility for the next four years. It is anticipated that nationally competitive grants and industry linkage will ensure the facility is sustainable in the longer term. The resource will be available to all Australian veterinary and allied researchers at all times through the application to the Executive Board, that meets every month and comprises the project manager and representatives of six of the contributing schools (one school contributes less funding than all of the others and so has agreed to attend only the annual meeting of the Board, in a non-voting capacity).

It is anticipated that researchers will access the coding portal via a web interface and create a focal dataset (e.g., all records related to cats with a diagnosis of chronic renal disease). They then further code and characterize the data (e.g., by answering questions related to the diagnostic techniques implemented, as evidenced by the free-text clinical records). These secondary data are then automatically saved to the VetCompass Australia database, and, when finished, the researcher can export this data interrogation with related animal and treatment fields to a spreadsheet for subsequent analysis.

Each veterinary school has existing or proposed studies that will interrogate VetCompass data. Some examples of projects already waiting for the resource are: the identification of spatial clustering of clinical cases of zoonotic infections; the development of predictive spatial models of clinical cases of zoonotic infections and identification of areas at most risk of clinical zoonotic infections; epidemiology and lifestyle factors associated with body condition score and diabetes; criteria-based and population-validated diagnostic systems for dermatology treatments; and the use of antimicrobials in the companion-animal sector and development of guidelines that instigate changes in usage to reduce the development of resistant organisms.

There are three main types of enquiry that the VetCompass facility invites: student, strategic, and major research projects. Student projects are intended chiefly for final-year veterinary student research projects and will typically focus on simple enquiries (e.g., breed-specific enquiries). These reports could describe the longevity and the demographics of the breed. For every condition to which a breed is predisposed, the students could report on the temporal and geographic distribution as well as risk factors and treatments. Several students may focus on one breed at the same time, and, for group projects, teams of students may code the same records. Strategic projects are designed for clinical academics, or early to mid-career researchers from the participating institutions. Importantly, strategic projects must not encroach directly on research questions that have been ear-marked for major research projects. The success of applications to VetCompass Australia does not depend on the track-record of the investigators. Instead, these projects are reserved for academic staff other than full professors and are capped at two concurrent projects per chief investigator.

Any researcher wanting to access the data held in the VetCompass system needs to apply to the Board, which uses set criteria to determine eligibility. The process followed depends on the type of project proposed. Student projects are allocated according to the local rules for each university, with the Board members from each institution selecting breeds in advance. Strategic project submissions are sought from each university then brought to the Board for ratification. Where the topic relates directly

to a current major project, the Chief Investigator on the relevant major project must also give approval. Ideas for major projects are submitted to the Board for review, who must agree on whether or not to advance a project. Such approval also depends on current commitments of the data to other projects.

There will be two calls for applications per year. Multi-year projects will be considered, but priority will be given to shorter projects with a defined outcome and likely output. Major research projects are expected to be led by senior academics within the consortium and with strong track records, since they are intended to be competitive for significant industry collaboration and national funding schemes. With a focus on quality over equity, these projects are largely dependent on the involvement of an epidemiologist as a key member of the investigatory team. Multi-institutional projects are preferred, but not at the compromise of quality.

### 3. Aims and Significance

The need for a nation-wide companion-animal disease-surveillance system to inform rational approaches to companion animal disease control was first triggered by a perceived rise in inherited disorders among companion animals [7–9]. Some of these are related to breed standards [10], while others are not [11]. Recent large-scale genomic investigations have clearly identified increased deleterious genetic variation in domestic dogs compared with ancestral canids, extending earlier estimations of multiple deleterious gene carriage [12]. VetCompass UK was developed as part of a response to the need for a surveillance system to describe and evaluate companion-animal disorders, and to formulate plans to eliminate them. Researchers have used the accumulated data related to nearly 6 million animals to publish articles that provide invaluable guidance to clinicians and breeders on breed predispositions and are informing the curricula of veterinary schools globally [13].

Epidemiological studies are powerful means of establishing correlations amongst different types of physical, spatial and temporal data to better understand many aspects of disease. For example, the use of routinely collected data by clinics or government departments facilitates an understanding of how patterns of infectious disease emerge. Outbreak data collected by state governments in Australia aided understanding of the 2007 equine influenza outbreak by allowing estimation of the relative contributions of network and spatial spread, results that will guide planning for control of future outbreaks [14,15]. There are numerous examples of clinical data being used to obtain insights into epidemiology and risk factors for small-animal diseases [16–20]. However, data from specialized clinics at veterinary schools are often used for such analyses where the case population is generally not representative of the general population for which inferences are required, thus resulting in selection bias. In contrast, VetCompass data are aggregated from primary care veterinary clinics, where the case population and case types better represent the general pet population, and therefore, research findings from such data have a greater internal and external validity.

Effective data linkage of clinical records for the development of spatial epidemiological tools is needed to quantify the geographical limits of infectious disease occurrence, the populations at risk, and to identify major risk factors for zoonotic infections of significance to public health. VetCompass Australia will provide the evidence to allow refinement of existing clinical protocols with the aim of implementing targeted control and mitigation strategies to directly decrease animal disease burden.

Companion-animal research is under-funded in Australia, with the chief sources of funding being confined to small non-profit organizations, pet-food manufacturers, and pharmaceutical manufacturers who, quite understandably, have partisan research needs. This can impede long-term, large-scale studies into the burden of disease, risk factors, and cost-effectiveness of treatments for companion animals. The merit of individual therapeutic agents may differ with chronicity and even breed, so it seems clearly beneficial to collect “big” data that may allow clinical guidelines for any disorders of companion animals to be fine-tuned as more evidence becomes available. Another good example concerns companion-animal vaccinations. VetCompass will enable researchers to investigate the extent to which Australian veterinarians follow international vaccination guidelines [21] and



improve surveillance and understanding of vaccine-related adverse reactions, which are thought to be under-reported.

Therapeutic guidelines are needed more than ever as the veterinary profession takes a leadership role in monitoring the use and abuse of antimicrobial agents that lead to antimicrobial resistance [22,23]. Microbial pathogens of companion animals cause disease in owners and allow antimicrobial resistance to emerge in humans. Methicillin-resistant *Staphylococcus aureus* (MRSA) is a recognized problem in equine hospitals and a recent study found that equine veterinarians had a very high rate of MRSA carriage [24,25]. MRSA has been transmitted between humans and domestic pet dogs and cats in both directions—zoonotic and reverse-zoonotic transmission [26]. MRSA costs Australian taxpayers between \$70 million and \$300 million in preventable human infections, and VetCompass will advance our understanding of the appropriateness of veterinary antimicrobial usage and the potential role of poor antimicrobial stewardship on resistance transference between pets and their owners.

The potential and power of natural language processing (NLP) to improve analysis of large or poorly coded datasets is considerable, and within the veterinary context this technology offers some particularly exciting possibilities when compared with human health initiatives. Existing health applications have tended to focus on either very small-scale datasets (e.g., hundreds of patients [27]) or publicly available data sources such as Twitter, with the incumbent issues of data provenance, and imprecise location data [28,29]. Emergent work on the application of NLP in the analysis of clinical notes for human patients and monitoring of disease outbreaks [30–32] is severely limited because patient privacy legislation limits access to these free-form notes. In contrast, access to high-quality de-identified free-form clinical notes within data in VetCompass will be relatively unfettered.

#### 4. The Relevance of VetCompass to the Needs of Companion-Animal Health

The paucity of comprehensive epidemiological data of clinical relevance is a problem facing almost all veterinary researchers in Australia, and the cost for any individual institution to address this gap is prohibitive. Recruiting veterinary clinics and their personnel to assist with active disease surveillance on a per-project basis is time-consuming and inefficient: it runs the risk of “respondent fatigue” among practitioners and progressive disinterest in data collection. The consortium has already had expressions of interest from more than 50 researchers at Australian universities seeking the types of data collected by VetCompass.

The project will support a wide range of research disciplines in key species and diverse breeds. There are enormous positive implications for the welfare of Australia’s companion animals and, ultimately, any animal for which individual records are kept, by providing the opportunity to evaluate the lifetime impact of disorders on the welfare of affected animals [33,34]. The prospects for longitudinal studies of individual animals are particularly exciting. Financial and policy gains for the public, industry, and government agencies will be made through research outcomes that promote cost-effective treatments, create better breeding guidelines, and document biosecurity risks.

The VetCompass Australia project will support a wide range of existing research projects as it provides both baseline data that can be augmented and an internally validated source of information. This is one possible means of ranking disorders and prioritizing prevention, eradication and management schemes, and research plans [35]. Prioritization can help funding bodies, such as the ARC and the Australian Government’s Department of Agriculture, to identify research priorities and effectively assess the value of investment in particular ventures. It will also assist the cat, dog, and horse breeders of Australia to decide how to prioritize inherited disorder disease mitigation programs and monitor the progress of these initiatives over time. This level of analysis is not currently possible and, at a national scale, could provide a significant return on investment.

#### 5. Relevance of VetCompass to the Needs of Veterinary Clinics

Evidence-based practice is the science of making informed, validated decisions for day-to-day clinical practice scenarios. This ideal is currently difficult to implement in the veterinary practice

context because of the limited availability of high-quality data. Unfortunately, some clinical decisions in veterinary practice are, of necessity, based on evidence generated in case studies, personal experience, extrapolations from human medicine, or advice from colleagues, which is considered the lowest level of evidence [36,37]. VetCompass will allow researchers in companion-animal medicine and surgery, as well as in the para-clinical sciences such as pathology, microbiology, virology, and parasitology, to benefit from the collective clinical experience of their colleagues in a national network of private practices and at other participating institutions. In aggregating these data, the Australian veterinary profession will, for the first time, generate a sound, broad-based dataset from which to make evidence-based recommendations for best practice in the diagnosis, treatment, and prevention of companion-animal diseases.

In Australia, there is almost no published information on age-, sex- and breed-based disorder prevalence and treatment for dogs. Equivalent data on cats and horses are even more limited. VetCompass offers the first opportunity in Australia to gather baseline demographic statistics (such as breed, age, sex, neuter status, insurance status, and bodyweight) on Australian dogs, cats, and horses. It will generate estimates of disorder prevalence for the most common disorders in the most common breeds and crossbreeds. This will identify risk factors for common disorders and reveal the most cost-effective treatments and their impact on animal survival and health. In addition, the data derived from this project may inform strategic breeding programs to improve the health and longevity of vulnerable breeds.

## **6. Use of VetCompass Data to Understand Patterns of Zoonotic and Emerging Infectious Diseases for Gaining Insight for Both Human and Animal Health**

It is anticipated that VetCompass data will become the primary source of information about temporal and spatial patterns in distribution of zoonotic and emerging infectious diseases in companion animals in Australia. This will not only enable early detection of diseases but also allow health authorities to monitor changes in disease incidence that will inform policies for their control. This will have positive implications for the health and wellbeing of both companion animals and humans.

Emerging infectious diseases in humans and wildlife populations are increasing worldwide, alongside other forms of global change [38–40]. A major challenge for science is to understand pathogens in terms of how they emerge, lie dormant, and sometimes disappear, and of how they drive aspects of population immunity. For conservation, production, and human health, there is an imperative to attain better data on the factors that influence the movement and transmission of disease from wildlife to humans. This affects, in part at least, the push for a One Health approach to the understanding of zoonoses [41–43].

VetCompass will provide a highly efficient and economical, passive and non-invasive system for monitoring and surveillance of companion-animal diseases of public health importance. Practice records will be uploaded daily automatically to the centralized repository. The real-time nature of the data means that information about the temporal and spatial distribution of diseases will be available as diseases occur. With disease modeling and further automation, VetCompass data will be useful for public health authorities and veterinary bodies by providing information on trends in disease occurrence, and predicting where and when diseases are likely to occur.

Through real-time surveillance of incoming data, the VetCompass platform could be instrumented to communicate alerts of emergent diseases, such as rabies or Hendra virus (e.g., if a cluster of veterinary practices starts to record neurological signs in multiple patients). The project also has the capacity to reveal how wildlife health may integrate with domestic animal and human health. Many veterinary clinics throughout Australia also attend to a large number of sick and injured wild animals during the course of their normal work, but most of the information that could be acquired from such accessions is not currently captured or analyzed by Wildlife Health Australia (WHA), which is the key group responsible for wildlife health surveillance in Australia ([www.wildlifehealth.com.au](http://www.wildlifehealth.com.au)). WHA currently uses sentinel clinic surveillance to capture such data and could clearly benefit from



wider sources, since summary reports on a selection of wildlife disease and mortality events are published quarterly through the National Animal Health Information System (NAHIS) in each issue of Animal Health Surveillance Quarterly (AHSQ). WHA also provides information to help fulfil Australia's reporting requirements to the World Organization for Animal Health (OIE). One caveat with wildlife data is that, in the absence of unique identifiers, such as microchips, one animal could theoretically be presented with the same disorder several times. Therefore, the prevalence data that emerge from wildlife would need to be interpreted with caution. Additionally, many practices do not create formal medical records for presented wildlife.

## 7. Use of VetCompass Data for Training Future Veterinary and Animal Scientists, and for Guiding Development of Veterinary Curricula

VetCompass data will be invaluable for training future veterinarians and animal scientists in epidemiology, biostatistics, clinical research and evidence-based practice. Most veterinary schools require veterinary and animal science students to undertake courses in research and enquiry, and to conduct research projects. For these students, as well as for full-time Honours or PhD research students, a national data-collection and mining platform offers a dynamic, high-quality and pre-funded dataset for self-contained research projects or modules of larger projects. The VetCompass project also has potential for facilitating the development of research careers of clinical academics with a substantial clinical service workload that reduces the time they have available to conduct primary research.

Much of the data harvested by VetCompass from electronic clinical records will be in plain text, providing myriad opportunities for NLP. Applications of NLP to clinical data have historically been marred by data-access restrictions and sparsity, but the volume of data collected by VetCompass is conducive to the application of pre-training of computational models, which has been shown to substantially boost accuracy [44,45]. There is significant value in the innovative NLP components of VetCompass. The consortium includes NLP scholars, whose work will boost the functionality of VetCompass by providing researchers with better search and coding tools, and improving clinicians' coding practices. This is a unique value proposition of the consortium in that research, and the infrastructure enabling that research, deliver mutual benefits.

The NLP components of the project span a broad spectrum of complexity. For example, among our plans to use NLP in the study of antimicrobial use, a central task will be the determination of the dosage instructions provided to the patient for a given prescription (e.g., "3 × 500 mg tablets daily"), as well as the classification of the diagnosis described in a given clinical note, e.g., via VeNom codes (an online database dictionary maintained by the Royal Veterinary College, see [www.venomcoding.org](http://www.venomcoding.org)). Methodologically, this is a relatively straightforward classification task but will rely on the availability of hand-coded VeNom codes from VetCompass UK and other sources to train supervised models, and almost certainly will involve extra data annotation. This will be critical in interpreting the context in which antimicrobials have been prescribed (or not prescribed), to aid in the collation of data on the use of antimicrobials in different clinical contexts. The NLP will go hand in hand with machine learning, in capturing and representing uncertainty in the dosage or diagnostic data, and aggregating these imprecise data at the level of species, breeds, clinicians, and practices.

The data-linkage opportunities in this project will also enable innovative research into the medical geography of veterinary diseases in Australia. This includes, for example, linkage with existing projects on the epidemiology of grass-seed foreign-body disease in dogs [46], and projects under development such as the University of Queensland's Spatial Epidemiology Laboratory ([www.spatialepilab.com](http://www.spatialepilab.com)), the University of Melbourne's temporo-spatial mapping of cat and dog snake envenomations (SnakeMap) and the University of Sydney's dog-care and management tracking tool, doglogbook ([www.doglogbook.com](http://www.doglogbook.com)). The measurable impact of this research will go beyond traditional research outputs. The VetCompass project will allow us to disseminate data and summary information at minimal or no charge to key stakeholders, including veterinarians, breeders, and potential pet purchasers. It is anticipated that feedback loops will permit the tailoring of breed-, sex- and age-specific

data so that differential diagnoses will be redefined according to probabilities in the light of these variables. Dissemination of this data to veterinary clinicians will allow for more individualized preventative health programs to be implemented.

In view of the anticipated benefits listed above, it is not surprising that VetCompass has the support of the Heads of all seven veterinary schools in Australia, and support from all the veterinary teaching hospitals based at universities, as they identify a genuine need for the advancement of research in their respective schools and clinics. Participating institutions have agreed to fund an on-going support position for a minimum of four years beyond the ARC infrastructure grant, so that researchers can maximize their interaction with the data over a long period. They are taking joint responsibility for funding, publication, and research output, in that each major contributor is given a seat on the executive board to participate in all governance issues.

The consortium is a comprehensive national group and consists of at least one chief investigator representative from each veterinary school in Australia, and is endorsed by Veterinary Schools of Australia and New Zealand (VSANZ), which was formed to promote strategic research alliances at the highest level of veterinary education. As with the pan-institutional group of veterinary hospital directors, VetCompass is one of a variety of initiatives supported by VSANZ to enhance collaboration across the sector.

The consortium has strong engagement (including co-appointments) with peak bodies such as the Australian Veterinary Association (AVA), the Australian and New Zealand College of Veterinary Scientists (ANZCVS), the Royal Veterinary College (RVC), the UK's Royal College of Veterinary Surgeons (RCVS), and the Royal Society for the Prevention of Cruelty to Animals (RSPCA) in both Australia and the UK. These representative bodies have given support to VetCompass both financially and in kind.

This project also brings unprecedented opportunities for sharing specialist expertise across institutions and increasing coverage of inter-disciplinary research across the sector. Negotiations that led to formation of the VetCompass consortium focused on equitable access to the resultant data and produced a governance structure (see below, 9. Project Governance) that ensures transparency and accountability.

## 8. Limitations and Challenges

The almost unparalleled scope and size of the dataset notwithstanding, the limitations imposed by the data collection method need careful consideration. Chief among these is the selection bias introduced by the collection of secondary data from animals presented at veterinary clinics rather than from the primary pet population. The availability and utilization of veterinary services is highly variable, as evidenced by the only moderate correlation between the spatial distribution of companion animals and veterinarians in Australia [47]. Spatial effects can be countered to some extent by continuing to broaden the base of data contributors in under-represented areas. However, this selection bias is of a much lower level than that in studies based on data from specialist clinics in veterinary schools or from selection of a convenience sample from the primary pet population. In addition, the data will be collected from a variety of veterinary practices, including rural, urban, and emergency clinics. Inclusion of a variety of clinic types and locations will help overcome the demographic and other inherent biases.

The other major considerations are those inherent to multi-source retrospective datasets, mainly arising from an inability to standardize the quality of data collected. The lack of uniformity in the diagnostic criteria, clinical management, and record-keeping across the large number of clinicians could result in misclassification and confounding biases. This conundrum is termed data-rich, information-poor (DRIP), but actions can be taken to ensure that the data can be mined appropriately [48]. Such measures include the application of robust scientific methods during the data cleansing, analysis, and interpretation phases. Additionally, there is some evidence that functional data-testing can overcome bad data or data-quality issues [49].

Another challenge is the limited skill of veterinary students, and veterinarians in general, to handle, manage, and analyze data, particularly big data generated by surveillance systems such as VetCompass. The primary focus of veterinary education is to train students in the treatment, prevention, and control of animal diseases, but there is limited focus on providing training in data management and analysis. Many other disciplines, such as business studies, have increased the emphasis on data analytics in response to big data from social media and businesses. It is anticipated that veterinary schools will need to keep pace and increase emphasis on training in data-handling and analysis in the veterinary curriculum, or offer specialized courses to interested students and veterinarians.

## 9. Project Governance

The shared facility is governed by an Executive Board, a steering sub-committee, and an advisory sub-committee. The Executive Board consists of senior representatives from six of the major contributing veterinary schools (one school elected to make a relatively minor financial contribution and accordingly agreed to step away from Board responsibilities from the outset). The Board holds monthly meetings to plan and review the VetCompass build-and-deployment progress. It is also responsible for reaching consensus on financial matters, personnel hiring, data access, citation and acknowledgment requirements. From time to time, the Board delegates tasks and responsibilities to selected groups from the Steering and Advisory committees for considering critical matters of eResearch infrastructure, data management, and deduplication of research topics. The ongoing nature of the VetCompass project means that, over time, participating veterinary schools may need to replace their nominated Board representative with another chief investigator. During the rollout and implementation of Phase 1, the steering sub-committee (comprising the ARC grant holders, the project manager, software engineers, and up to 10 volunteers invited by the Board from the Australian research community) reviews and advises on functionality and use of VetCompass. The non-executive advisory sub-committee advises the Board on areas of expertise, and provides ongoing outreach through publication, conference papers, social media, research supervision, and scholarly user support for VetCompass. It consists of Board-appointed project collaborators and specialist advisors to advise the Board on matters relating to the building and deployment of VetCompass. It includes an NLP specialist and representatives of the RVC, the RSPCA Australia, and the AVA. Additional members may be appointed at the Board's discretion.

The VetCompass platform comprises two main data systems: the Backend Database and the Analysis Portal. With the provided LIEF funding, both systems are being developed to maximize access for Australian researchers. The portal is available 24/7 and will support concurrent access. The backend database collects data records from participating veterinary clinics and stores them in a hosted SQL database with password-protected access. Access to this database is restricted to the collection manager (the Project Manager of the LIEF grant) and Executive Board members. The system is currently designed to host 200 million animal records, which is estimated to be sufficient until 2020. Federation of the backend database with the UK and other international data is a key outcome of the current government funding. De-identified and aggregated data will be available to Australian researchers through the analysis portal, designed to facilitate research projects and provide searching, aggregation, downloading, and coding services. Access to VetCompass requires authentication of approved researchers. Further improvement to the existing portal is a core aspect of the project.

VetCompass in the UK has been highly productive with the publication of peer-reviewed mortality studies on dogs [3] and cats [4]; longevity and prevalence reports on several breeds (e.g., Cavalier King Charles spaniels [50], pugs [51] and German Shepherd dogs [52]); multi-breed assessments of gastric dilation-volvulus [53], dystocia [54], diabetes mellitus [55], and audits of anti-microbial usage in companion animal practice [56]. The Australian consortium plans to apply the same methods to data on local dogs and publish matching reports so that comparison between breeds in the two countries can be made.

It is anticipated that VetCompass in Australia will continue to grow in unison with its sister project in the UK and that, together, a family of similar projects using the same approach to electronic veterinary patient records will be established throughout the developed world.

Access is mediated by a chief investigator of a given researcher's institution and the Project Manager. These gatekeepers will also actively monitor proposed research questions to minimize the duplication of research. The data-collection process includes de-identification to comply with Australian privacy laws.

## 10. Conclusions

VetCompass Australia has brought together all seven of the veterinary schools in Australia and is the first Australia-wide surveillance system that collects real-time records from veterinary clinics and aggregates them for researchers to monitor and record companion-animal diseases and their treatments. It delivers to Australian researchers sustainable and cost-effective access to authoritative data from hundreds of practitioners. These clinical records will reveal longitudinal trends in the prevalence of inherited and acquired diseases and their optimal treatments, revolutionise veterinary clinical auditing in Australia, help the veterinary profession to rank research priorities, and assure research-led curricula in veterinary schools.

**Acknowledgments:** The authors would like to thank colleagues at the Royal Veterinary College, notably Noel Kennedy, David Church, and Dan O'Neill. The support and assistance of the Australian Veterinary Association, the heads of the Australian veterinary schools, the Australian and New Zealand College of Veterinary Scientists, the Australian Research Council, and the RSPCA Australia are all warmly acknowledged.

**Author Contributions:** Paul McGreevy conceived the collection of veterinary electronic patient records for surveillance purposes and Paul McGreevy and Jeremy Hammond conceived the current national consortium; David Brodbelt contributed details on interface and database structure; Timothy Baldwin contributed details on natural language processing; Paul McGreevy, Peter Thomson, Navneet K. Dhand, David Raubenheimer, Sophie Masters, Caroline S. Mansfield, Timothy Baldwin, Ricardo J. Soares Magalhaes, Jacquie Rand, Peter Hill, Anne Peaston, James Gilkerson, Martin Combs, Shane Raidal, Peter Irwin, Peter Irons, Richard Squires, David Brodbelt and Jeremy Hammond wrote the paper.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Animal Medicines Australia. *Pet Ownership in Australia 2016 Report*; Animal Medicines Australia Pty Ltd.: Barton, Australia, 2016.
2. Australian Companion Animal Council. *Contribution of the Pet Care Industry to the Australian Economy*, 7th ed.; Rockwell Communications Inc.: East Kew, Victoria, Australia, 2010.
3. O'Neill, D.G.; Church, D.B.; McGreevy, P.D.; Thomson, P.C.; Brodbelt, D.C. Prevalence of disorders recorded in dogs attending primary-care veterinary practices in England. *PLoS ONE* **2014**, *9*, e90501. [[CrossRef](#)] [[PubMed](#)]
4. O'Neill, D.G.; Church, D.B.; McGreevy, P.D.; Thomson, P.C.; Brodbelt, D.C. Prevalence of disorders recorded in cats attending primary-care veterinary practices in England. *Vet. J.* **2014**, *202*, 286–291. [[CrossRef](#)] [[PubMed](#)]
5. Ward, M.P.; Kelman, M. Companion animal disease surveillance: A new solution to an old problem? *Spat. Spatio-Temporal Epidemiol.* **2011**, *2*, 147–157. [[CrossRef](#)] [[PubMed](#)]
6. Guernier, V.; Milinovich, G.J.; Bezerra Santos, M.A.; Haworth, M.; Coleman, G.; Soares Magalhaes, R.J. Use of big data in the surveillance of veterinary diseases: Early detection of tick paralysis in companion animals. *Parasite Vector* **2016**, *9*, 303. [[CrossRef](#)] [[PubMed](#)]
7. McGreevy, P.D. Breeding for Quality of Life. *Anim. Welf.* **2007**, *16*, 125–128.
8. McGreevy, P.D.; Nicholas, F. Some practical solutions to welfare problems in dog breeding. *Anim. Welf.* **1999**, *8*, 329–341.
9. Nicholas, F.W.; Arnott, E.R.; McGreevy, P.D. Does hybrid vigour have any utility in dog breeding? *Vet. J.* **2016**, *214*, 77–83. [[CrossRef](#)] [[PubMed](#)]

10. Asher, L.; Summers, J.F.; Diesel, G.; McGreevy, P.D.; Collins, L.M. Inherited defects in pedigree dogs I: Disorders that are related to breed standards. *Vet. J.* **2009**, *182*, 402–411. [[CrossRef](#)] [[PubMed](#)]
11. Summers, J.F.; Diesel, G.; Asher, L.; McGreevy, P.D.; Collins, L.M. Inherited defects in pedigree dogs II: Disorders that are not related to breed standards. *Vet. J.* **2010**, *183*, 39–45. [[CrossRef](#)] [[PubMed](#)]
12. Marsden, C.D.; Ortega-Del Vecchyo, D.; O'Brien, D.C.; Taylor, J.F.; Ramirez, O.; Vilà, C.; Marques-Bonet, T.; Schnabel, R.D.; Wayne, R.K.; Lohmueller, K.E. Bottlenecks and selective sweeps during domestication have increased deleterious genetic variation in dogs. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 152–157. [[CrossRef](#)] [[PubMed](#)]
13. Degeling, C.; Fawcett, A.; Collins, T.; Hazel, S.; Johnson, J.; Lloyd, J.; Phillips, C.; Stafford, K.; Tzioumis, V.; McGreevy, P. Students' opinions on welfare and ethics issues for companion animals in Australian and New Zealand veterinary schools. *Aust. Vet. J.* **2016**, *95*, 189–193. [[CrossRef](#)] [[PubMed](#)]
14. Firestone, S.M.; Ward, M.P.; Christley, R.M.; Dhand, N.K. The importance of location in contact networks: Describing early epidemic spread using spatial social network analysis. *Prev. Vet. Med.* **2011**, *102*, 185–195. [[CrossRef](#)] [[PubMed](#)]
15. Firestone, S.M.; Christley, R.M.; Ward, M.P.; Dhand, N.K. Adding the spatial dimension to the social network analysis of an epidemic: Investigation of the 2007 outbreak of equine influenza in Australia. *Prev. Vet. Med.* **2012**, *106*, 123–135. [[CrossRef](#)] [[PubMed](#)]
16. Barmettler, R.; Schweighauser, A.; Bigler, S.; Grooters, A.M.; Francey, T. Assessment of exposure to *Leptospira serovars* in veterinary staff and dog owners in contact with infected dogs. *J. Am. Vet. Med. Assoc.* **2011**, *238*, 183–188. [[CrossRef](#)] [[PubMed](#)]
17. Mellanby, R.J.; Handel, I.G.; Clements, D.N.; Bronsvort, B.M.; Lengeling, A.; Schoeman, J.P. Breed and sex risk factors for canine babesiosis in South Africa. *J. Vet. Intern. Med.* **2011**, *25*, 1186–1189. [[CrossRef](#)] [[PubMed](#)]
18. Liem, B.P.; Dhand, N.K.; Pepper, A.E.; Barrs, V.R.; Beatty, J.A. Clinical findings and survival in cats naturally infected with feline immunodeficiency virus. *J. Vet. Intern. Med.* **2013**, *27*, 798–805. [[CrossRef](#)] [[PubMed](#)]
19. Rubin, J.A.; Suran, J.N.; Brown, D.C.; Agnello, K.A. Factors associated with pathological fractures in dogs with appendicular primary bone neoplasia: 84 Cases (2007–2013). *J. Am. Vet. Med. Assoc.* **2015**, *247*, 917–923. [[CrossRef](#)] [[PubMed](#)]
20. Kohler, I.; Ballhausen, B.D.; Stockhaus, C.; Hartmann, K.; Wehner, A. Prevalence of and risk factors for feline hyperthyroidism among a clinic population in Southern Germany. *Tierarztl. Prax. Ausg. K Kleintiere/Heimtiere* **2016**, *44*, 149–157. [[CrossRef](#)] [[PubMed](#)]
21. Day, M.J.; Horzinek, M.C.; Schultz, R.D.; Squires, R.A. Vaccination Guidelines Group (VGG) of the World Small Animal Veterinary Association (WSAVA). WSAVA Guidelines for the vaccination of dogs and cats. *J. Small Anim. Pract.* **2016**, *57*, 4–8. [[CrossRef](#)] [[PubMed](#)]
22. Heller, J.; Kelly, L.; Reid, S.W.; Mellor, D.J. Qualitative risk assessment of the acquisition of Meticillin-resistant *Staphylococcus aureus* in pet dogs. *Risk Anal.* **2010**, *30*, 458–472. [[CrossRef](#)] [[PubMed](#)]
23. Heller, J.; Innocent, G.T.; Denwood, M.; Reid, S.W.; Kelly, L.; Mellor, D.J. Assessing the probability of acquisition of Meticillin-resistant *Staphylococcus aureus* (MRSA) in a dog using a nested stochastic simulation model and logistic regression sensitivity analysis. *Prev. Vet. Med.* **2011**, *99*, 211–224. [[CrossRef](#)] [[PubMed](#)]
24. Cuny, C.; Witte, W. MRSA in equine hospitals and its significance for infections in humans. *Vet. Microbiol.* **2016**, *200*, 59–64. [[CrossRef](#)] [[PubMed](#)]
25. Jordan, D.; Simon, J.; Fury, S.; Moss, S.; Giffard, P.; Maiwald, M.; Southwell, P.; Barton, M.D.; Axon, J.E.; Morris, S.G.; et al. Carriage of Methicillin-resistant *Staphylococcus aureus* by veterinarians in Australia. *Aust. Vet. J.* **2011**, *89*, 152–159. [[CrossRef](#)] [[PubMed](#)]
26. Weese, J.S.; Dick, H.; Willey, B.M.; McGeer, A.; Kreiswirth, B.N.; Innis, B.; Low, D.E. Suspected transmission of Methicillin-resistant *Staphylococcus aureus* between domestic pets and humans in veterinary clinics and in the household. *Vet. Microbiol.* **2006**, *115*, 148–155. [[CrossRef](#)] [[PubMed](#)]
27. Martinez, D.; Ananda-Rajah, M.R.; Suominen, H.; Slavin, M.A.; Thursky, K.A.; Cavedon, L. Automatic detection of patients with invasive fungal disease from free-text computed tomography (CT) scans. *J. Biomed. Inform.* **2014**, *53*, 251–260. [[CrossRef](#)] [[PubMed](#)]
28. Culotta, A. Towards detecting influenza epidemics by analyzing Twitter messages. In Proceedings of the First Workshop on Social Media Analytics, Washington, DC, USA, 25–28 July 2010; ACM: New York, NY, USA, 2010; pp. 115–122.



29. Signorini, A.; Segre, A.M.; Polgreen, P.M. The use of Twitter to track levels of disease activity and public concern in the US during the influenza A H1N1 pandemic. *PLoS ONE* **2011**, *6*, e19467. [[CrossRef](#)] [[PubMed](#)]
30. Heinze, D.T.; Morsch, M.L.; Holbrook, J. *Mining Free-Text Medical Records*; American Medical Informatics Association: Bethesda, MA, USA, 2001.
31. Jensen, P.B.; Jensen, L.J.; Brunak, S. Mining electronic health records: Towards better research applications and clinical care. *Nat. Rev. Genet.* **2012**, *13*, 395–405. [[CrossRef](#)] [[PubMed](#)]
32. LePendou, P.; Iyer, S.V.; Bauer-Mehren, A.; Harpaz, R.; Mortensen, J.M.; Podchiyska, T.; Ferris, T.A.; Shah, N.H. Pharmacovigilance using clinical notes. *Clin. Pharmacol. Ther.* **2013**, *93*, 547–555. [[CrossRef](#)] [[PubMed](#)]
33. Collins, L.M.; Asher, L.; Summers, J.F.; Diesel, G.; McGreevy, P.D. Welfare epidemiology as a tool to assessing the welfare impact of inherited defects on the pedigree dog population. *Anim. Welf.* **2010**, *19*, 67–75.
34. Collins, L.M.; Asher, L.; Summers, J.; McGreevy, P. Getting priorities straight: Risk assessment and decision-making in the improvement of inherited disorders in pedigree dogs. *Vet. J.* **2011**, *189*, 147–154. [[CrossRef](#)] [[PubMed](#)]
35. McGreevy, P.D.; Bennett, P.B. Challenges and paradoxes in the companion animal niche. Darwinian selection, selective breeding and the welfare of animals. *Anim. Welf.* **2010**, *19*, 11–16.
36. Ebell, M.H.; Siwek, J.; Weiss, B.D.; Woolf, S.H.; Susman, J.; Ewigman, B.; Bowman, M. Strength of recommendation taxonomy (SORT): A patient-centered approach to grading evidence in the medical literature. *Am. Fam. Physician* **2004**, *69*, 548–556. [[CrossRef](#)] [[PubMed](#)]
37. Antman, E.M.; Lau, J.; Kupelnick, B.; Mosteller, F.; Chalmers, T.C. A comparison of results of meta-analyses of randomized control trials and recommendations of clinical experts: Treatments for myocardial infarction. *JAMA* **1992**, *268*, 240–248. [[CrossRef](#)] [[PubMed](#)]
38. Jones, K.E.; Patel, N.G.; Levy, M.A.; Storeygard, A.; Balk, D.; Gittleman, J.L.; Daszak, P. Global trends in emerging infectious diseases. *Nature* **2008**, *451*, 990–993. [[CrossRef](#)] [[PubMed](#)]
39. Jones, B.A.; Grace, D.; Kock, R.; Alonso, S.; Rushton, J.; Said, M.Y.; McKeever, D.; Mutua, F.; Young, J.; McDermott, J.; et al. Zoonosis emergence linked to agricultural intensification and environmental change. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 8399–8404. [[CrossRef](#)] [[PubMed](#)]
40. Olson, S.H.; Benedum, C.M.; Mekaru, S.R.; Preston, N.D.; Mazet, J.A.; Joly, D.O.; Brownstein, J.S. Drivers of Emerging Infectious Disease Events as a Framework for Digital Detection. *Emerg. Infect. Dis.* **2015**, *21*, 1285–1292. [[CrossRef](#)] [[PubMed](#)]
41. Daszak, P.; Cunningham, A.A.; Hyatt, A.D. Emerging infectious diseases of wildlife—Threats to biodiversity and human health. *Science* **2000**, *287*, 443–449. [[CrossRef](#)] [[PubMed](#)]
42. Rabinowitz, P.; Conti, L. One Health and emerging infectious diseases: Clinical perspectives. *Curr. Top. Microbiol.* **2013**, *365*, 17–29.
43. Tompkins, D.M.; Carver, S.; Jones, M.E.; Krkosek, M.; Skerratt, L.F. Emerging infectious diseases of wildlife: A critical perspective. *Trends Parasitol.* **2015**, *31*, 149–159. [[CrossRef](#)] [[PubMed](#)]
44. Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G.S.; Dean, J. Distributed representations of words and phrases and their compositionality. In Proceedings of the 26th International Conference on Neural Information Processing Systems, Lake Tahoe, NV, USA, 5–10 December 2013; pp. 3111–3119.
45. Qu, L.; Ferraro, G.; Zhou, L.; Hou, W.; Schneider, N.; Baldwin, T. Big Data Small Data, in Domain Out-of-Domain, Known Word Unknown Word: The Impact of Word Representations on Sequence Labelling Tasks. In Proceedings of the Nineteenth Conference on Computational Natural Language Learning, Beijing, China, 30–31 July 2015; pp. 83–93.
46. Hicks, A.; Golland, D.; Heller, J.; Malik, R.; Combs, M. Epidemiological investigation of grass seed foreign body-related disease in dogs of the Riverina District of rural Australia. *Aust. Vet. J.* **2016**, *94*, 67–75. [[CrossRef](#)] [[PubMed](#)]
47. Smyth, G.B.; East, I.J.; Wicks, R.M. Spatial and temporal evaluation of veterinarians and veterinary employers relative to human and domesticated animal populations in Australia 2002–2012. *Aust. Vet. J.* **2015**, *93*, 137–144. [[CrossRef](#)] [[PubMed](#)]
48. Tien, J.M. Big Data: Unleashing Information. *J. Syst. Sci. Syst. Eng.* **2013**, *22*, 127–151. [[CrossRef](#)]
49. Gudipati, M.; Rao, S.; Naju, D.; Mohan, N.D.; Gajja, N.K. Big Data: Testing Approach to Overcome Quality Challenges. *Infosys Labs Brief.* **2013**, *11*, 65–74.



50. Summers, J.F.; O'Neill, D.G.; Church, D.B.; Thomson, P.C.; McGreevy, P.D.; Brodbelt, D.C. Prevalence of disorders recorded in Cavalier King Charles Spaniels attending primary-care veterinary practices in England. *Canine Genet. Epidemiol.* **2015**, *2*, 4. [[CrossRef](#)] [[PubMed](#)]
51. O'Neill, D.G.; Darwent, E.C.; Church, D.B.; Brodbelt, D.C. Demography and health of Pugs under primary veterinary care in England. *Canine Genet. Epidemiol.* **2016**, *3*, 5. [[CrossRef](#)] [[PubMed](#)]
52. O'Neill, D.G.; Coulson, N.R.; Church, D.B.; Brodbelt, D.C. Demography and disorders of German Shepherd Dogs under primary veterinary care in the UK. *Canine Genet. Epidemiol.* **2017**, *4*, 7. [[CrossRef](#)] [[PubMed](#)]
53. O'Neill, D.G.; Case, J.; Boag, A.K.; Church, D.B.; McGreevy, P.D.; Thomson, P.C.; Brodbelt, D.C. Epidemiology of gastric dilation-volvulus in dogs attending emergency-care veterinary practices in the UK. *J. Small Anim. Pract.* **2017**. [[CrossRef](#)] [[PubMed](#)]
54. O'Neill, D.G.; O'Sullivan, A.M.; Manson, E.A.; Church, D.B.; Boag, A.K.; McGreevy, P.D.; Brodbelt, D.C. Canine dystocia in 50 UK first-opinion emergency-care veterinary practices: Prevalence and risk factors. *Vet. Rec.* **2017**. [[CrossRef](#)] [[PubMed](#)]
55. Mattin, M.; O'Neill, D.; Church, D.; McGreevy, P.D.; Thomson, P.C.; Brodbelt, D. An epidemiological study of diabetes mellitus in dogs attending first opinion practice in the UK. *Vet. Rec.* **2014**, *174*, 349. [[CrossRef](#)] [[PubMed](#)]
56. Buckland, E.L.; O'Neill, D.; Summers, J.; Mateus, A.; Church, D.; Redmond, L.; Brodbelt, D. Characterisation of antimicrobial usage in cats and dogs attending UK primary care companion animal veterinary practices. *Vet. Rec.* **2016**, *179*, 489. [[CrossRef](#)] [[PubMed](#)]



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).